Analysis-driven optimisation of Pandora

Isobel Mawby Sim/Reco WG 14/03/22

Reconstruction → Analysis

Pandora pattern recognition → Particle characterisation → Nue/numu selection → Energy estimation → Delta CP

CM Sept 2021 LBL Session
https://indico.fnal.gov/event/46504/contributions/223982/attachments/147395/188851/CM_LBL_21_08_21.pdf

CM Sept 2021 DUNE FD Sim/Reco Session
https://indico.fnal.gov/event/46504/contributions/224069/attachments/147572/189113/CM_SR_23_09_21.pdf

LBL WG Meeting Nov 2021
https://indico.fnal.gov/event/51697/contributions/227258/attachments/148827/191354/LBL_01_11_21.pdf

CM Jan 2022 LBL Session
https://indico.fnal.gov/event/50215/contributions/232787/attachments/151288/195431/CM_LBL_26_01_22.pdf

A Traditional Selection Procedure

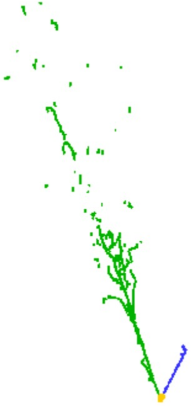
- Credit to **Dom Brailsford** (Lancaster University) for initial development and continued support and discussion – thank you!
- Our approach differs to the CVN as our input is the information of the reconstructed candidate **leading lepton only** (should it exist)

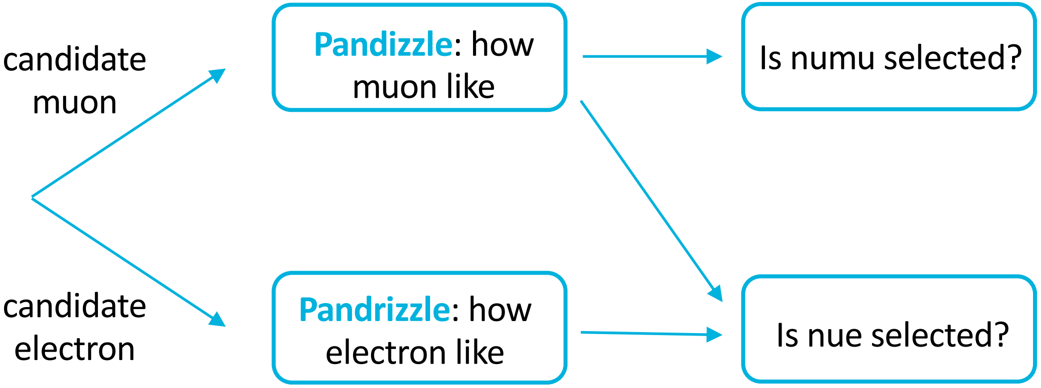
Pandrizzle*: a BDTG that assigns a score to a shower in the interval is [-1, 1] reflecting how electron-like it is

Pandizzle*: a BDTG that assigns a score to a track in the interval [-1, 1] reflecting how muon-like it is

*credit to **Dom Brailsford** for development

Reduce down to candidate muon/electron





candidate muon

candidate electron

Pandizzle: how muon like

Pandrizzle: how electron like

Is numu selected?

Is nue selected?

Selection details in BACKUP

Evaluation Metrics: CP Sensitivity

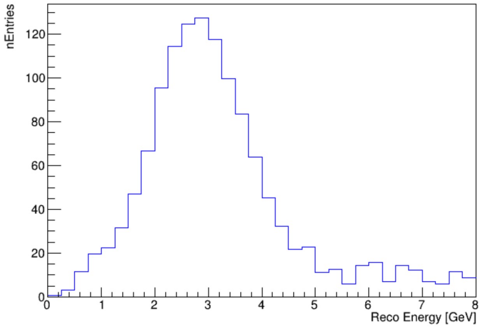
repeat for # delta CP values

repeat for # events

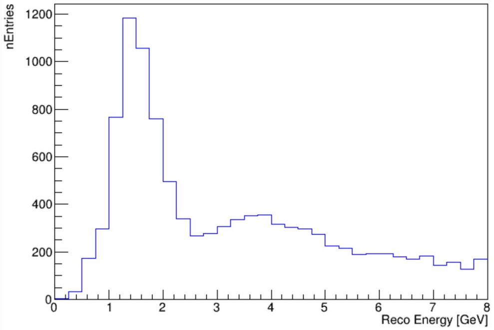
- Pass event through reconstruction chain and perform selection

Is CC nue selected? → no → Is CC numu selected?

yes

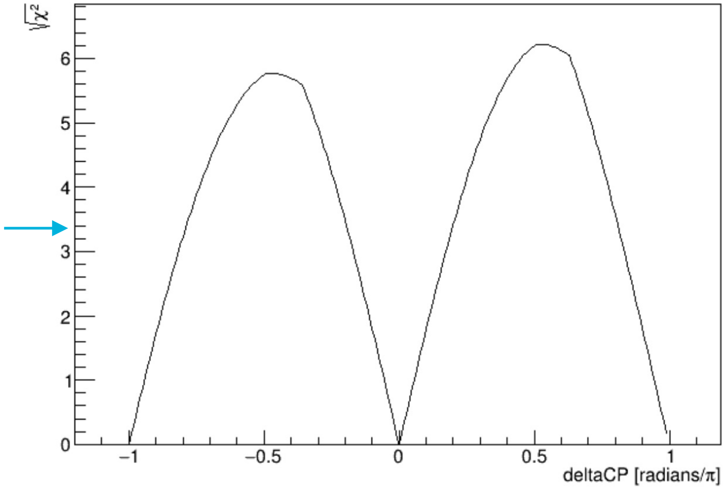


yes



- Fit reconstructed energy spectra to a CP conserving hypothesis

$$\sqrt{\chi^2} = \min\left(\sqrt{\chi^2}_{\delta_{CP}=0}, \sqrt{\chi^2}_{\delta_{CP}=\pi}\right),\ \chi^2 = 2\sum_i\left(\underset{\delta_{CP}=0,\pi}{\underset{\uparrow}{\boldsymbol{E_i}}} - O_i + \underset{\delta_{CP}\neq 0,\pi}{\underset{\uparrow}{\boldsymbol{O_i}}} \log\frac{O_i}{E_i}\right)$$



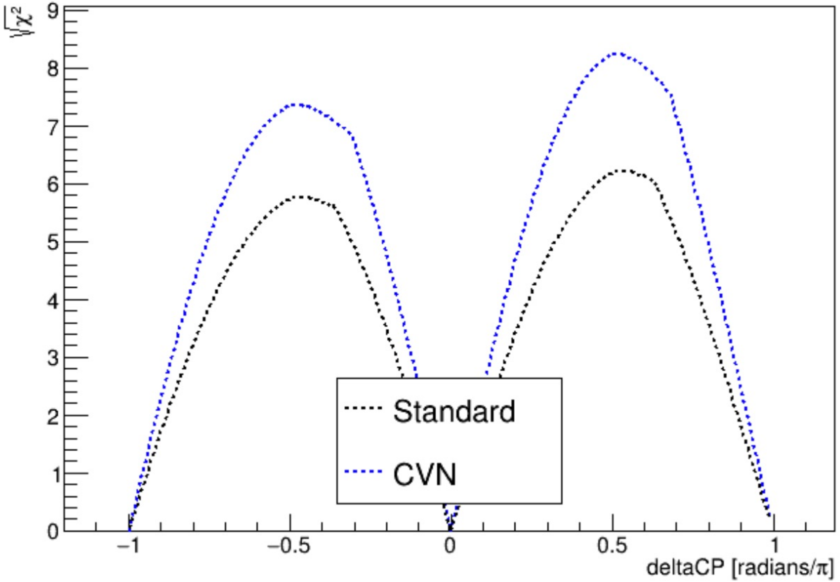
Very basic plot:

- **No statistical** or **systematic** uncertainties
- All other oscillation parameters **fixed** to their central NuFit 4.0 values in fit
- But **suits purpose** as 'quick' feedback to reconstruction changes

Benchmarking

- As previously mentioned our resolution plots are not quite ready, so let's focus on our **CPV metric**
- The CVN is very, very good – consequence of the superior **nue selection efficiency**

	Nue Efficiency	Nue Purity	Nue BG
CVN	82.7%	90.9%	99.6%
'izzle selection'	60.0%	67.1%	98.6%

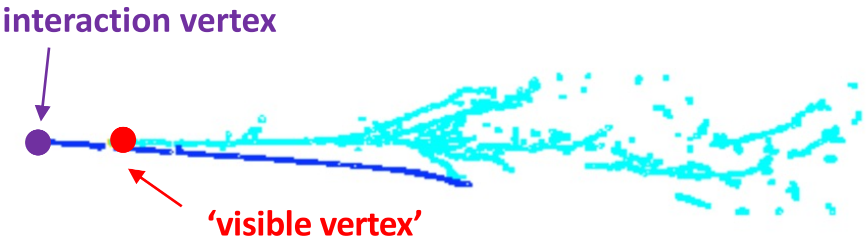


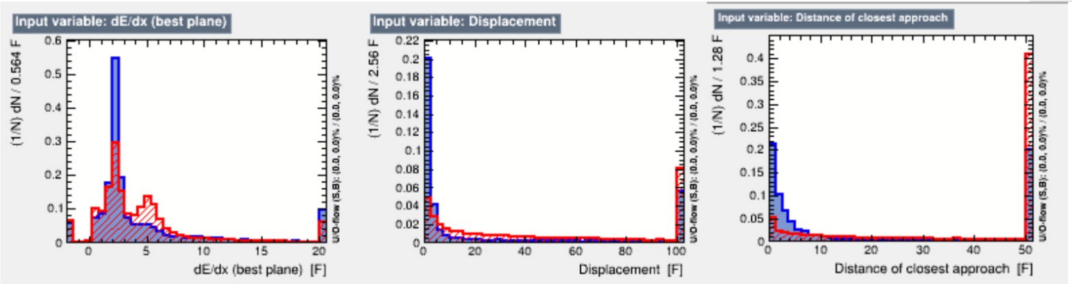
- We've been performing iterative **cheating studies*** starting with **very broad** investigations and using the results to identify **specific areas** to improve

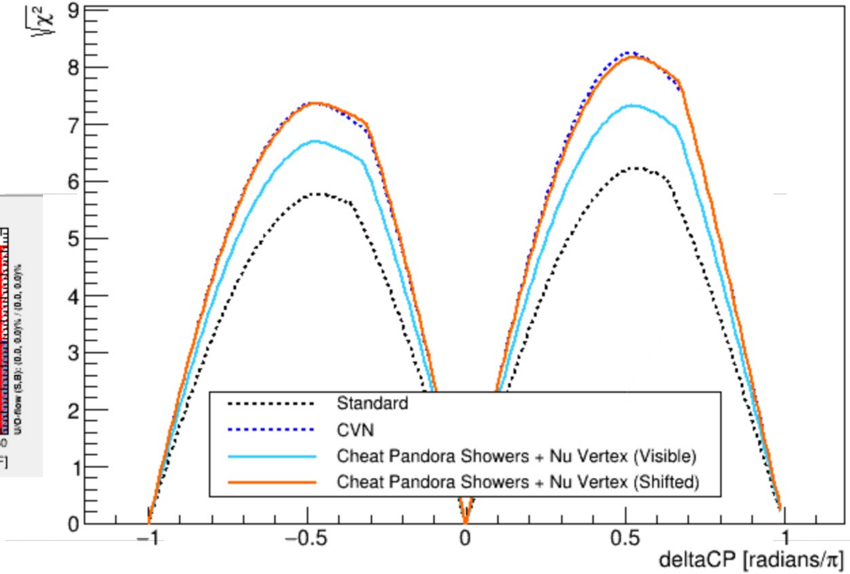
*** Cheating Studies:** When working with MC data we can slot in the truth for a particular reconstruction task i.e. pretend that a specific reconstruction task was performed perfectly

Leading Reconstruction Failure

- To cut a very long story short, we find that **the shower reconstruction drives the sensitivity improvements** and that the position of the **shower vertex is incredibly important**
- These findings suggest that it's the **initial shower region that's important here**, and our physics supports this – we're ultimately improving the **electron/gamma separation** in Pandrizzle

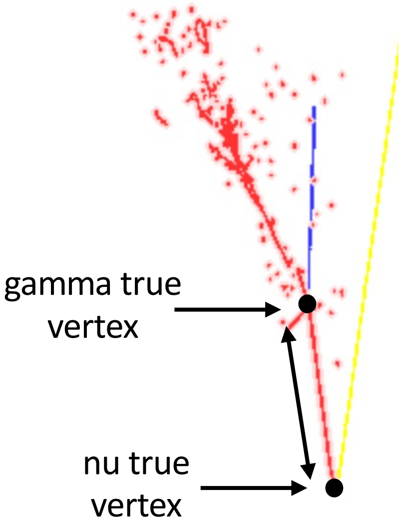


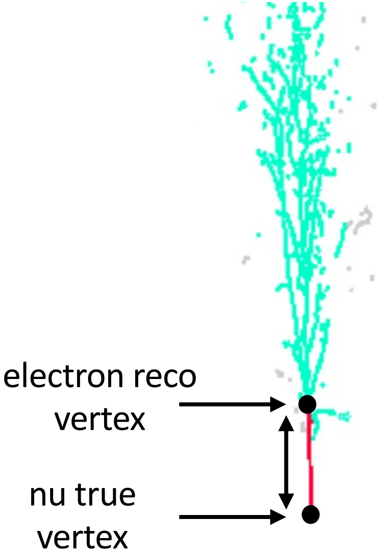


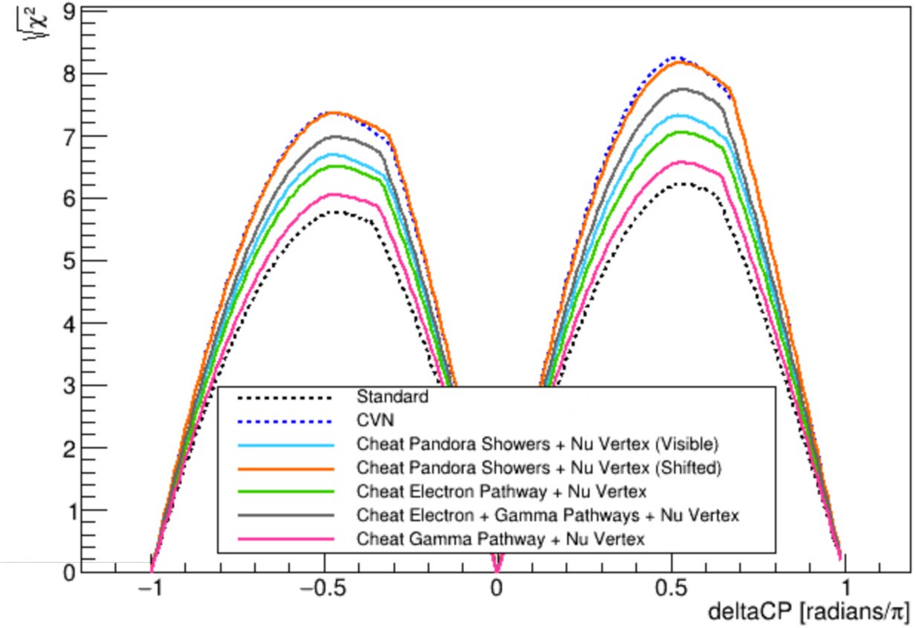


Cheating the Initial Shower Region

- To investigate this further lets create some cheating algorithms that, in the final stages of Pandora:
    - **truncate** gamma pfos by removing all non-gamma hits in the gamma true vertex → true nu vertex region
    - **extend** electron pfos by bringing in all true hits in the reco → true electron vertex region
- Also cheat the neutrino vertex but allow the rest of the reconstruction to **continue as normal**







From Cheating to Real Reconstruction

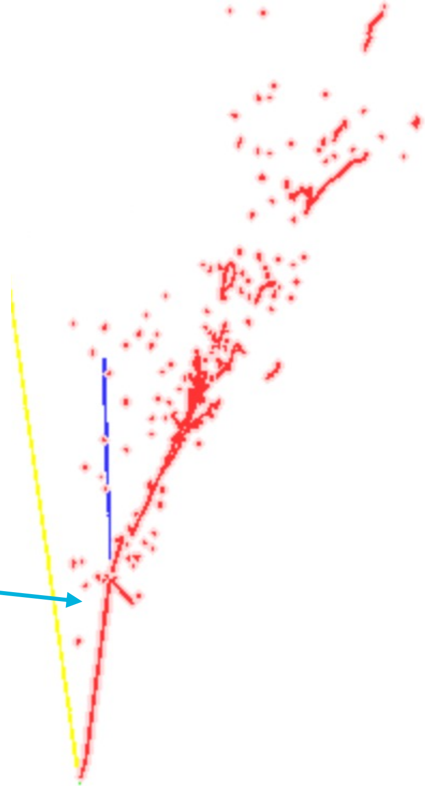
- We have now found a **specific and well defined reconstruction failure** that, if fixed, has been shown to return large gains in the sensitivity

- So, let's develop the reconstruction to fix this!

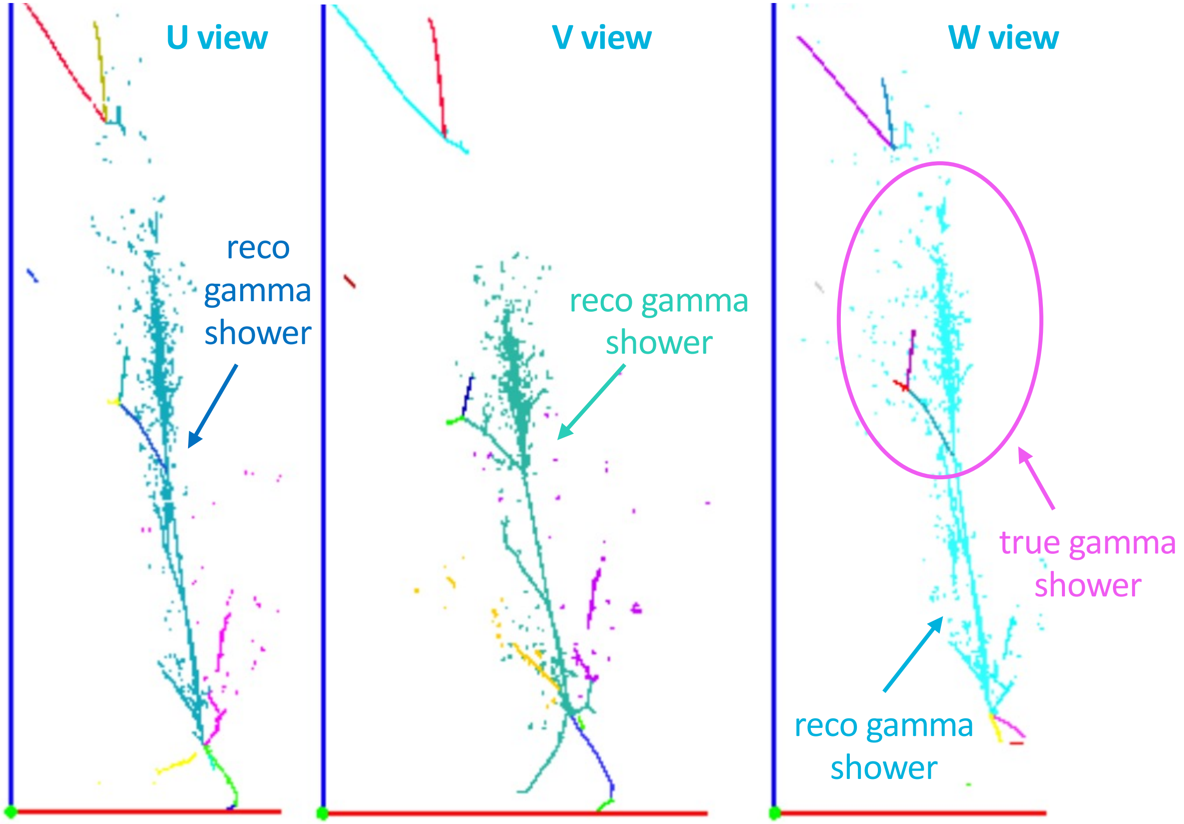
- First, we need to develop a method to find the **pathways** that

  - **electrons should have** taken to get back to the neutrino vertex

  - **gammas have mistakenly** taken to get back to the neutrino vertex



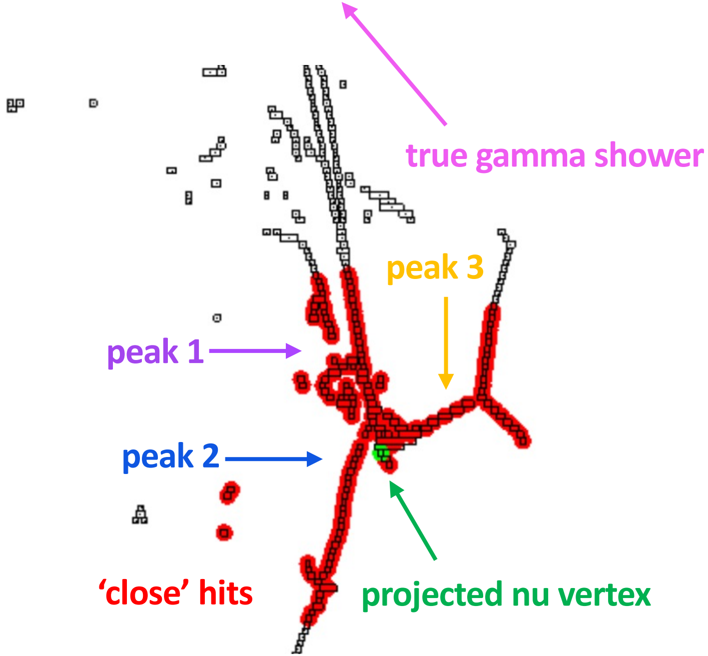
Finding the Connection Pathway(s)

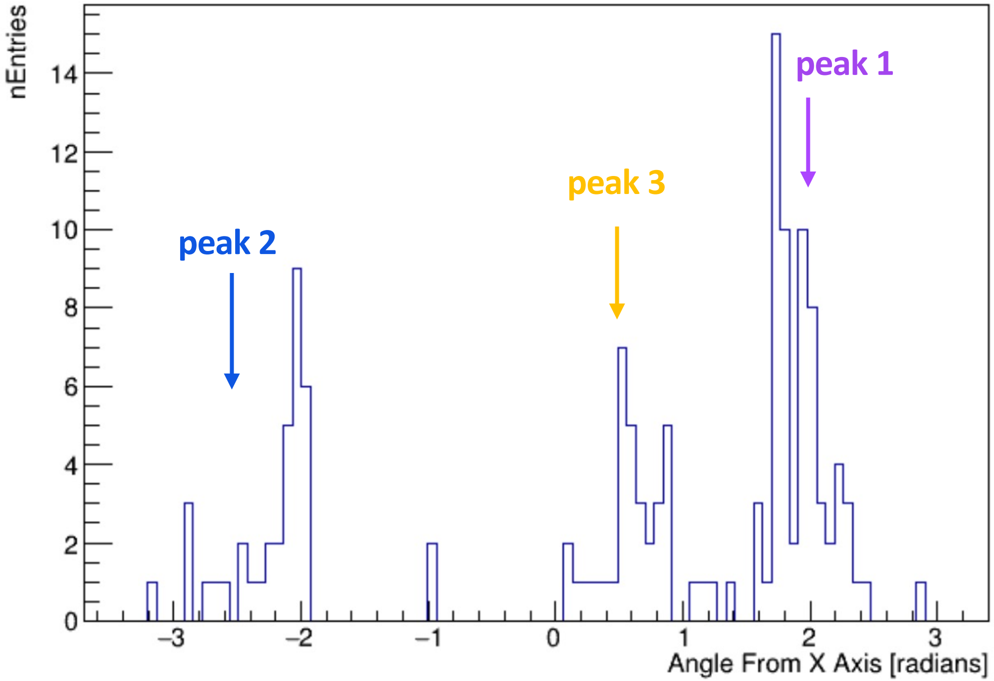


- Let’s go through the mechanics of this, with the following example
- A true photon has made its way **back to the neutrino vertex** by merging **with several tracks** coming out of the neutrino vertex
- It’s a good example because it demonstrates how **complex** the pfos are that we’re trying to fix

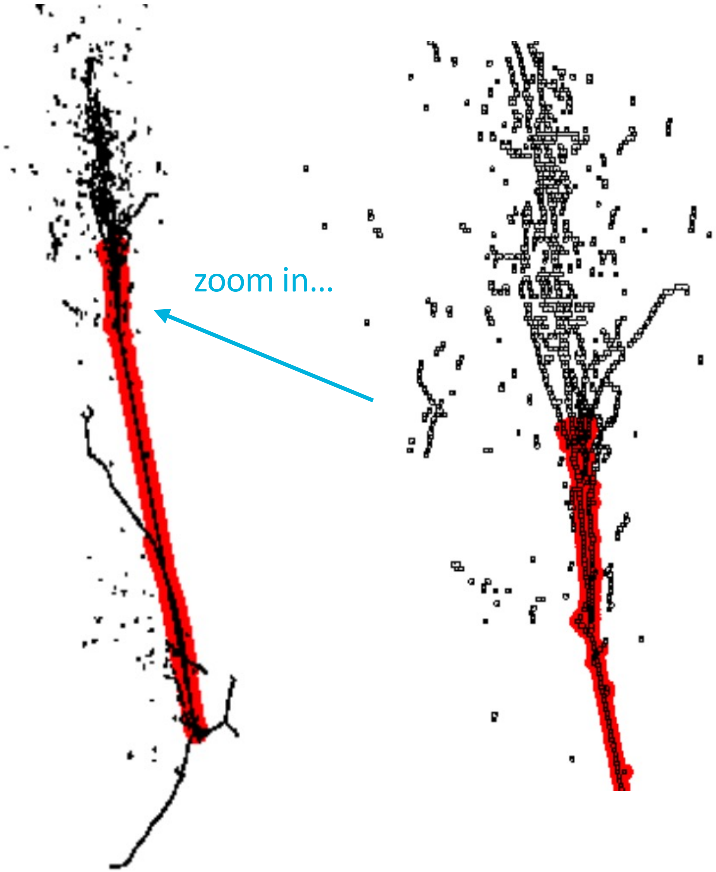
Finding the Connection Pathway

- First, we need to identify the directions that our brains follow **out from the neutrino vertex to get to the shower that's under our investigation**
- We therefore, plot **the angular distribution of hits** within a 'close region' of our neutrino vertex



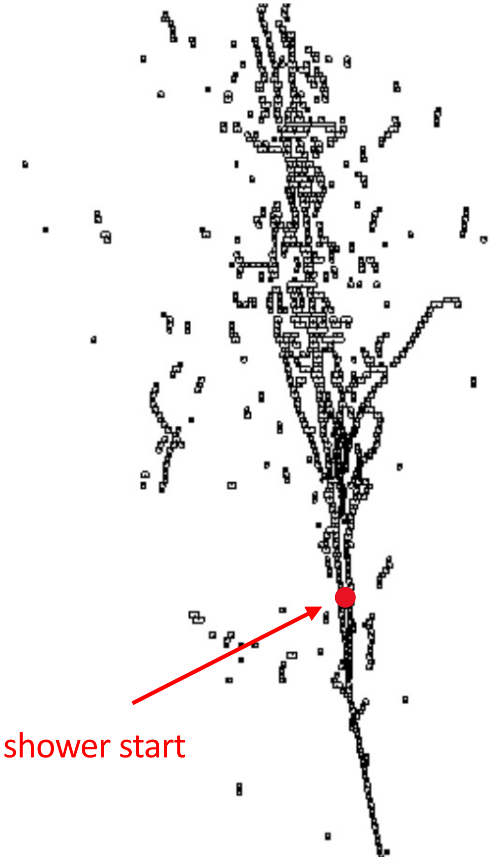


Finding the Connection Pathway

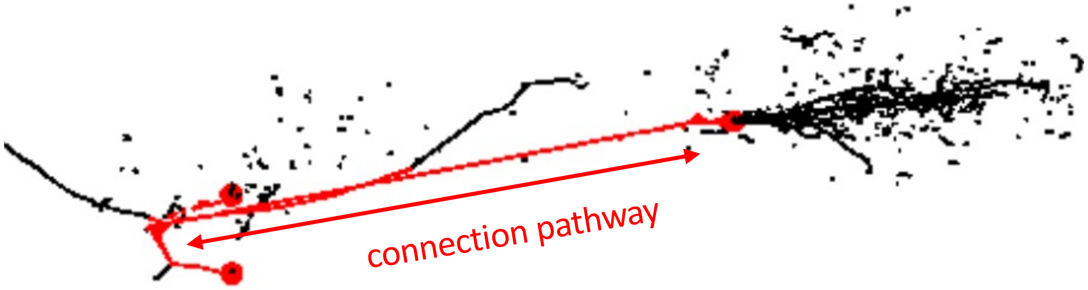


- We next **walk along each pathway** collecting the hits we intercept as we go along
- After each step, a **running fit** is performed to determine the direction of our next step – this allows us to **follow bends** in the pathway

Finding the Connection Pathway



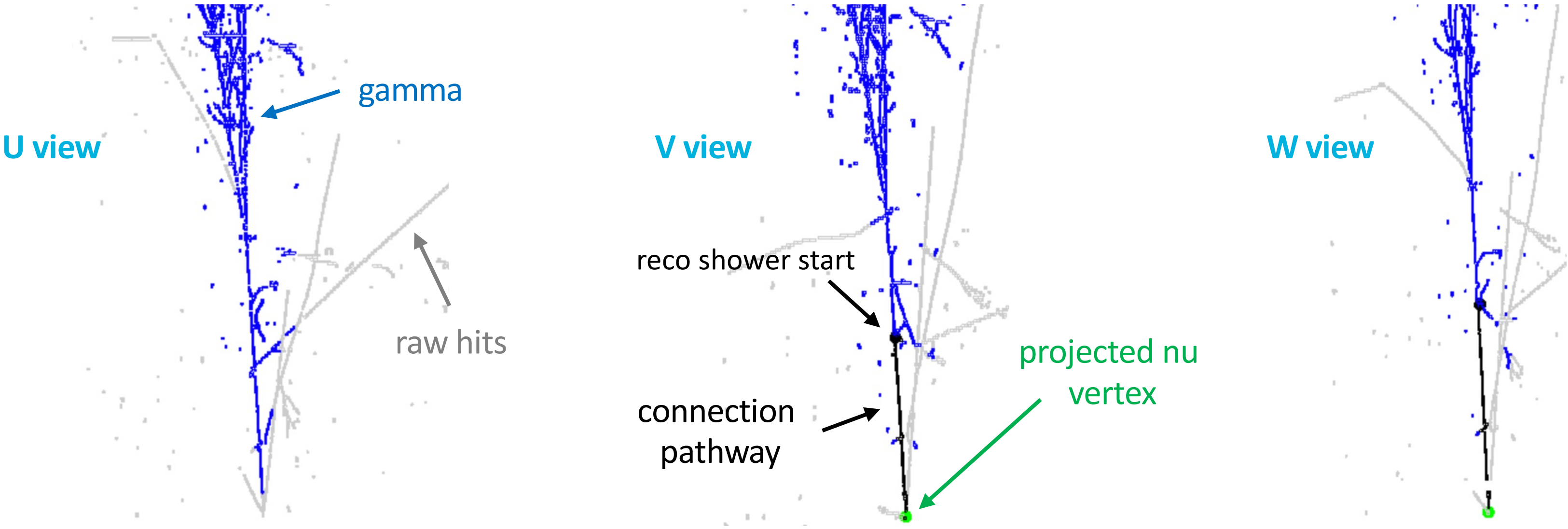
- Once our potential shower spine is found, we then use **calorimetric and topological information** to identify the position of our shower start
- Our connecting pathway is defined as the collected hits in the **region between the shower start and the neutrino vertex**



Gamma Truncation Algorithm

- With the connecting pathway mechanics we can now develop a reconstruction algorithm to **truncate the gamma showers**
- We loop through our **reconstructed showers** and for each:

1) Find the connecting pathways in **each 2D view** that allow the potential gamma pfo to travel back to the neutrino vertex



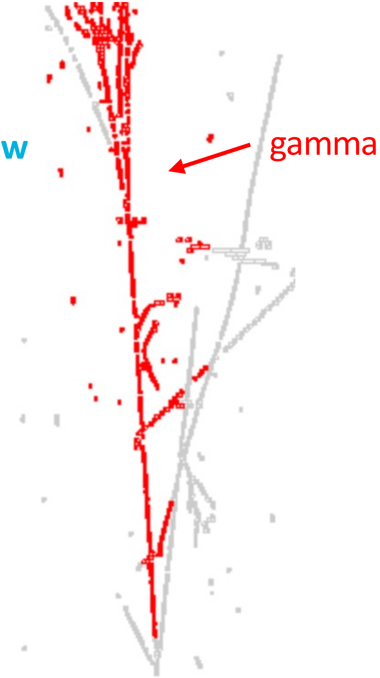
2) The same mistakes will not be made in each 2D view, so find the 2D connecting pathway that would **(but have not)** allow(ed) the gamma to travel back to the neutrino vertex in that view and match between views to obtain **3D connecting pathways**

Gamma Truncation Algorithm

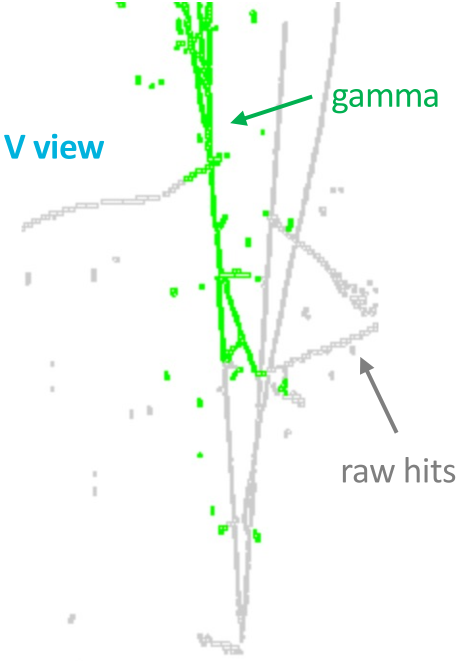
3) Assess each connecting pathway on whether it is **a true gamma pathway**

4) If it is not, remove the hits of the connecting pathway

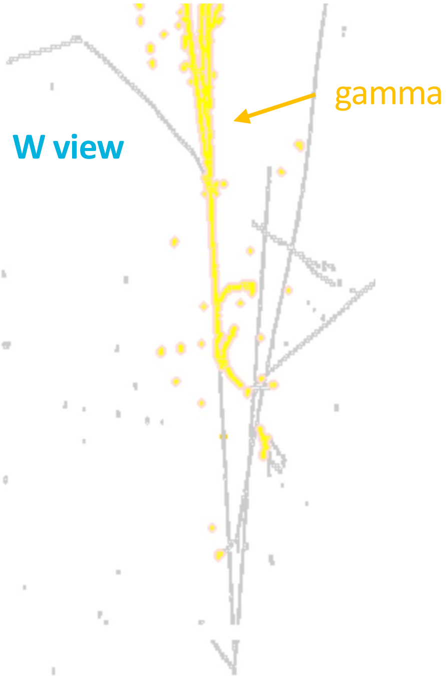
U view



V view

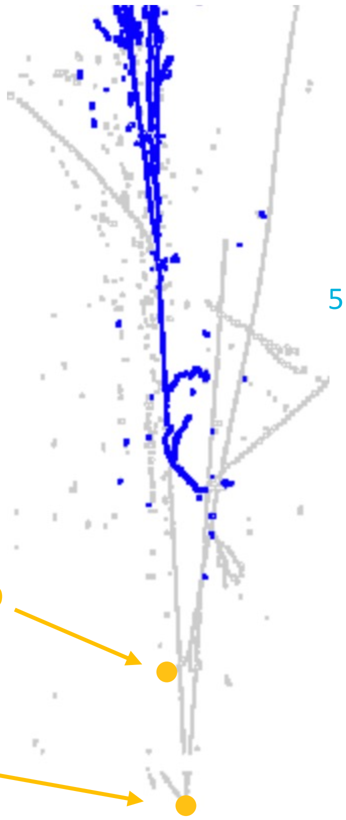


W view



Gamma Truncation Algorithm

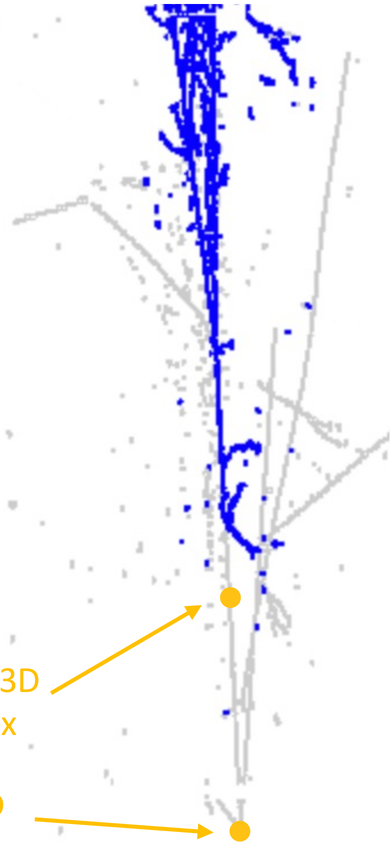
Sometimes the remnant 3D hits can result in a **poor placement** of the reconstructed gamma vertex that **cancels out our work in removing the connection pathway**



projected reconstructed 3D gamma vertex

projected 3D nu vertex

5) If a 3D connection pathway was found, **combine** the three 2D shower start vertices to obtain **a 3D shower start vertex**



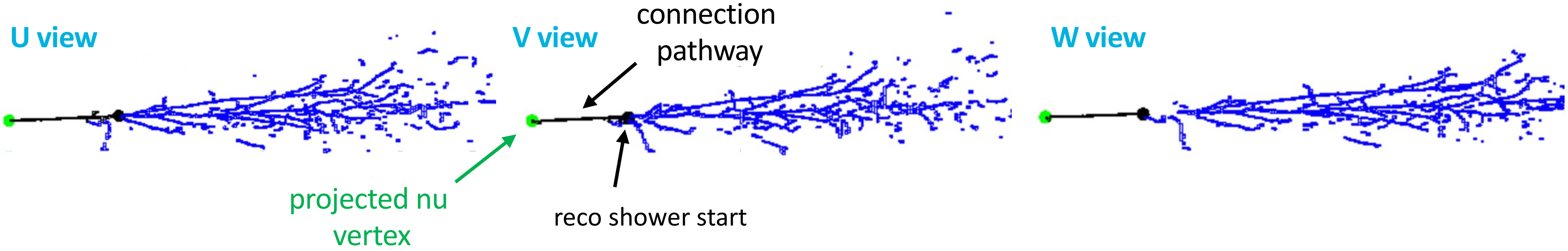
projected reconstructed 3D gamma vertex

projected 3D nu vertex

Electron Extension Algorithm

- Analogously, we can use the connection pathway mechanics in a reconstruction algorithm to **extend the electron showers**
- We loop through our **reconstructed showers** and for each:

1) The connecting pathways of the potential electron pfo to the neutrino vertex are found in **each 2D view**

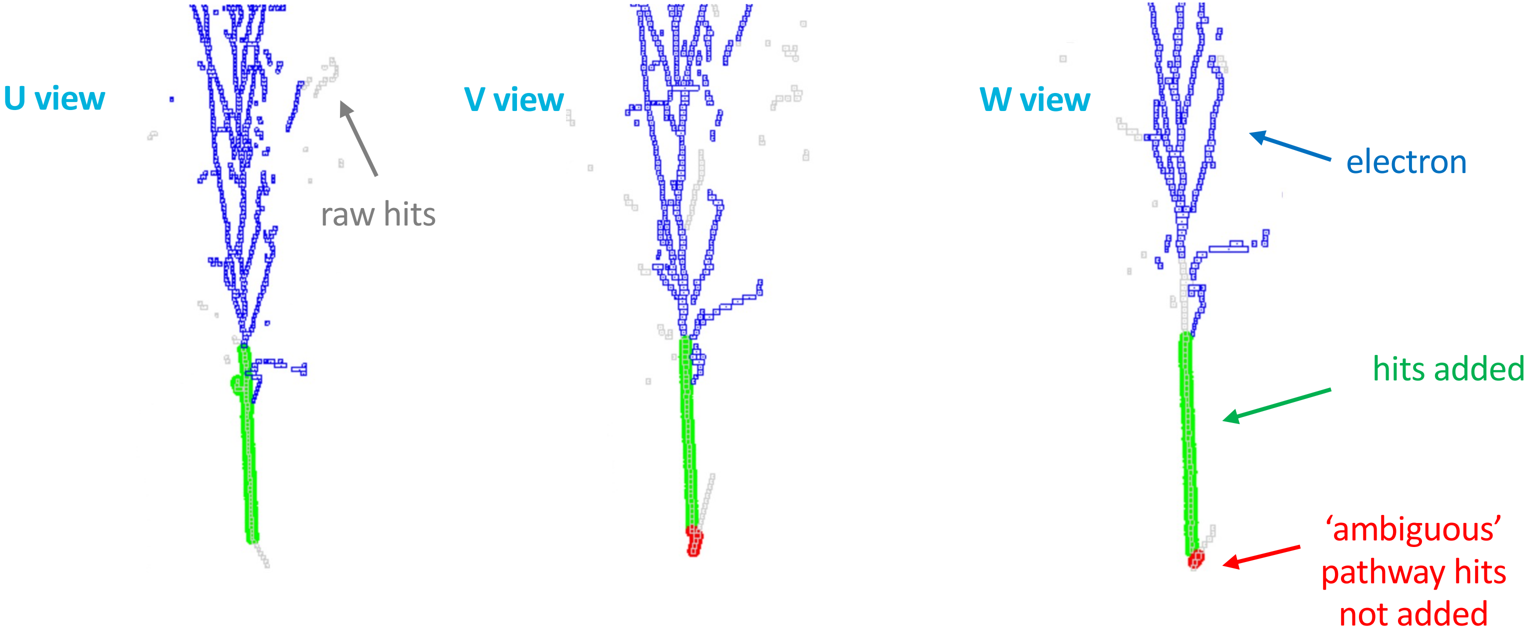


2) If the potential electron pfo is truly an electron it **will have a 3D connecting pathway** – it is therefore required that a 2D connecting pathway is found in each 2D view **and** that these pathways are **consistent in 3D**

Electron Extension Algorithm

3) The connecting pathway is then assessed on whether it is a **true electron pathway**

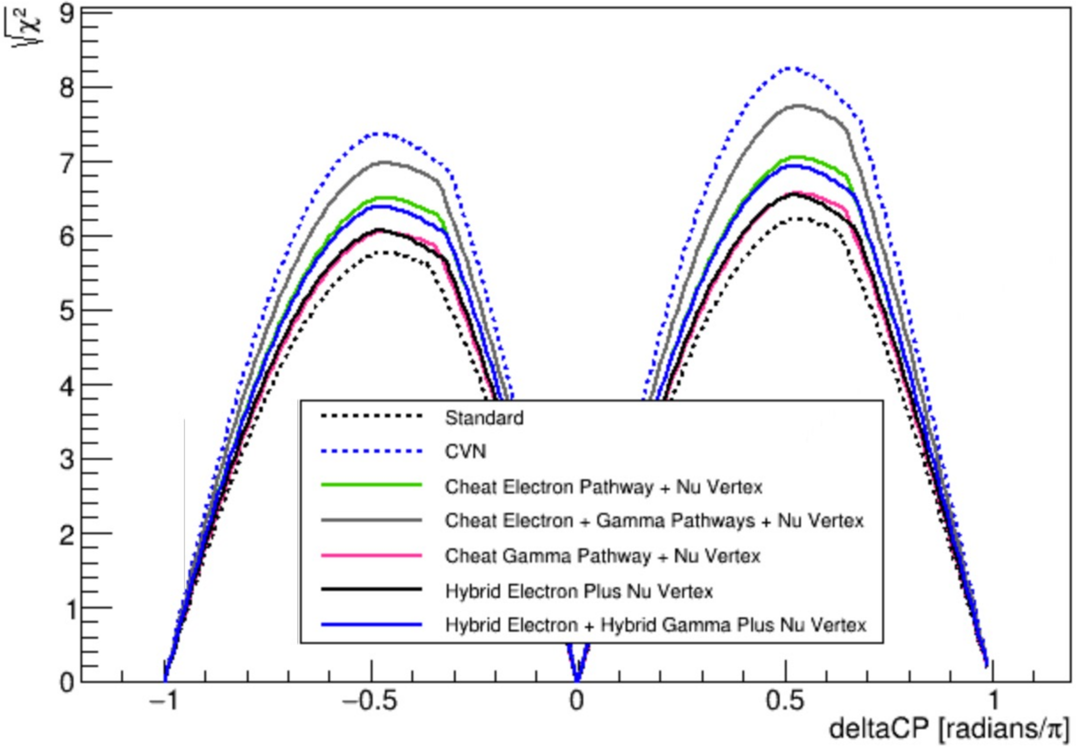
4) If it is, the hits of the connecting pathway are added into the pfo but with **caution so as not to contaminate the dEdx of the electron**



6) For future use, **the identification of the pfo as an electron is saved**

‘Hybrid Algorithms’

- At the moment, the nature of the connecting pathway is **cheated** by looking at **the true identity of the pfo under investigation**
- This allowed us to set the reconstruction mechanics and to validate their performance
- First version performance **is looking good!**
- To bring the hybrid config closer to the full shower cheat config
  - Improvements **have been made** to the hybrid electron and gamma algorithms
  - Remove **‘low hit pfos’** from training and selection
  - Use the **‘is electron’** information to set the Pandrizzle BDT **displacement variable to zero** and the dedx **variable to an electron-like value**



* The neutrino vertex is still being cheated..

From Hybrid to Real Reconstruction

- We're still relying on **two cheats**
  - **Cheating the neutrino vertex**
    - This can be removed by developing an **alternative vertexing procedure** for events with high energy showers (not sure if this will be included in this body of work)
  - **Cheating the connecting pathway addition/removal decision**
    - This will be removed by the development of a **'connecting pathway' BDT** which could use variables such as
      - the **length** of the connecting pathway
      - the **dedx stability**
      - the **topological agreement** of the connecting pathway and shower region
      - etc…

Conclusion

- The leading limiting reconstruction failure wrt CPV is the **reconstruction of the initial region of showers**
- A **hybrid electron extension tool** and **hybrid gamma truncation tool** has been created and the performance of a 'version 1' understood
- Work is now focused on making the **connection pathway assessment a real reconstruction decision** by the development of a **BDT**
- Following this, we plan to focus on the improvements we can extract from the **selection** itself
  - **Additional BDT variables**? Extracted from the connection pathway?
  - Using the fact that the pfo was extended (or contracted) and therefore has been **previously thought** to be an electron (gamma)
  - Enforcing a **'low hit'** pfo cut in the training and selection

BACKUP

Is this a nue event?

Pandizzle → muons
Pandrizzle → electrons

Reject if the **reconstructed** neutrino vertex is outside the DUNE fiducial volume

↓

Assign all showers **pandrizzle** scores, choosing the **pandrizzliest** shower to be the electron candidate

↓

Reject event if the candidate electron pandrizzle score falls **below** a cut value

To increase purity, remove numu background

Assign all tracks **pandizzle** scores, choosing the **pandizzliest** track to be the muon candidate

↓

Reject event if the candidate muon pandizzle score falls **above** a cut value

Tune pandrizzle cuts such that the **deltaCP sensitivity coverage** is optimised

Is this a numu event?

Reject if the **reconstructed** neutrino vertex is outside the DUNE fiducial volume

↓

Assign all tracks **pandizzle** scores, choosing the **pandizzliest** track to be the muon candidate

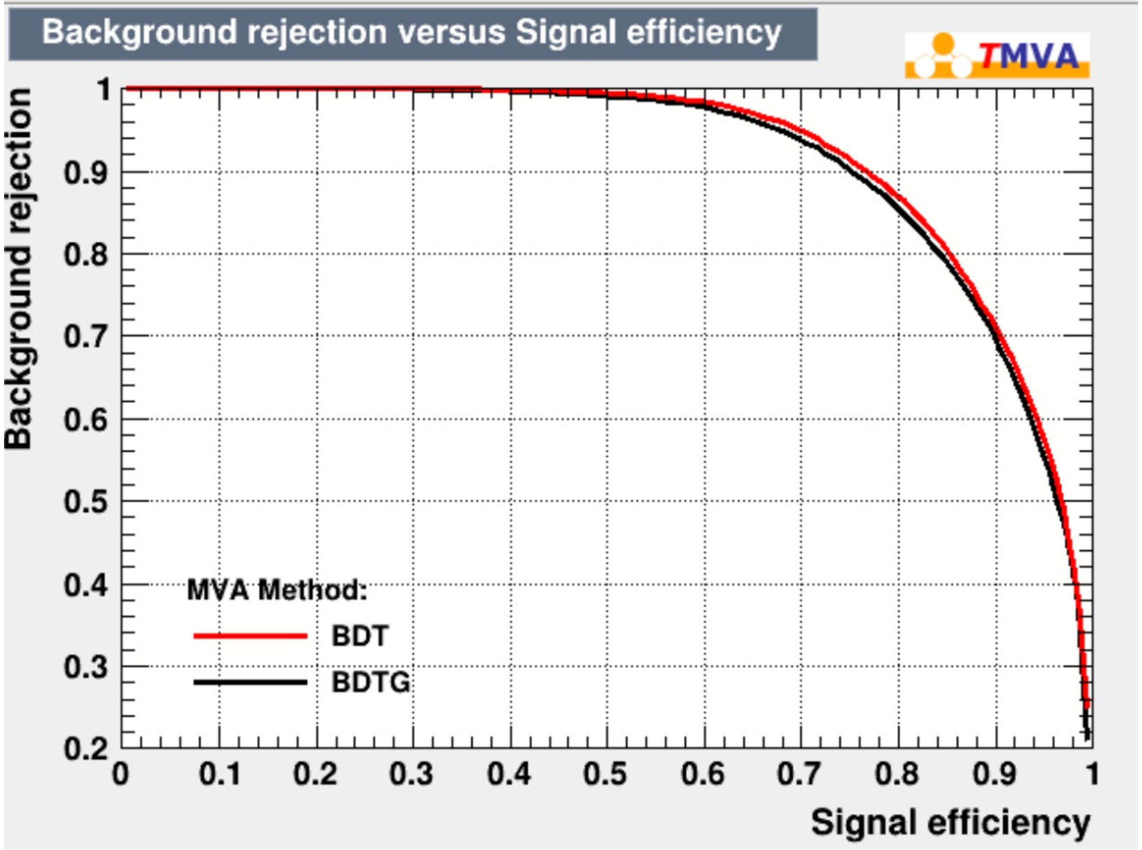
↓

Reject event if the candidate muon pandizzle score falls **below** a cut value

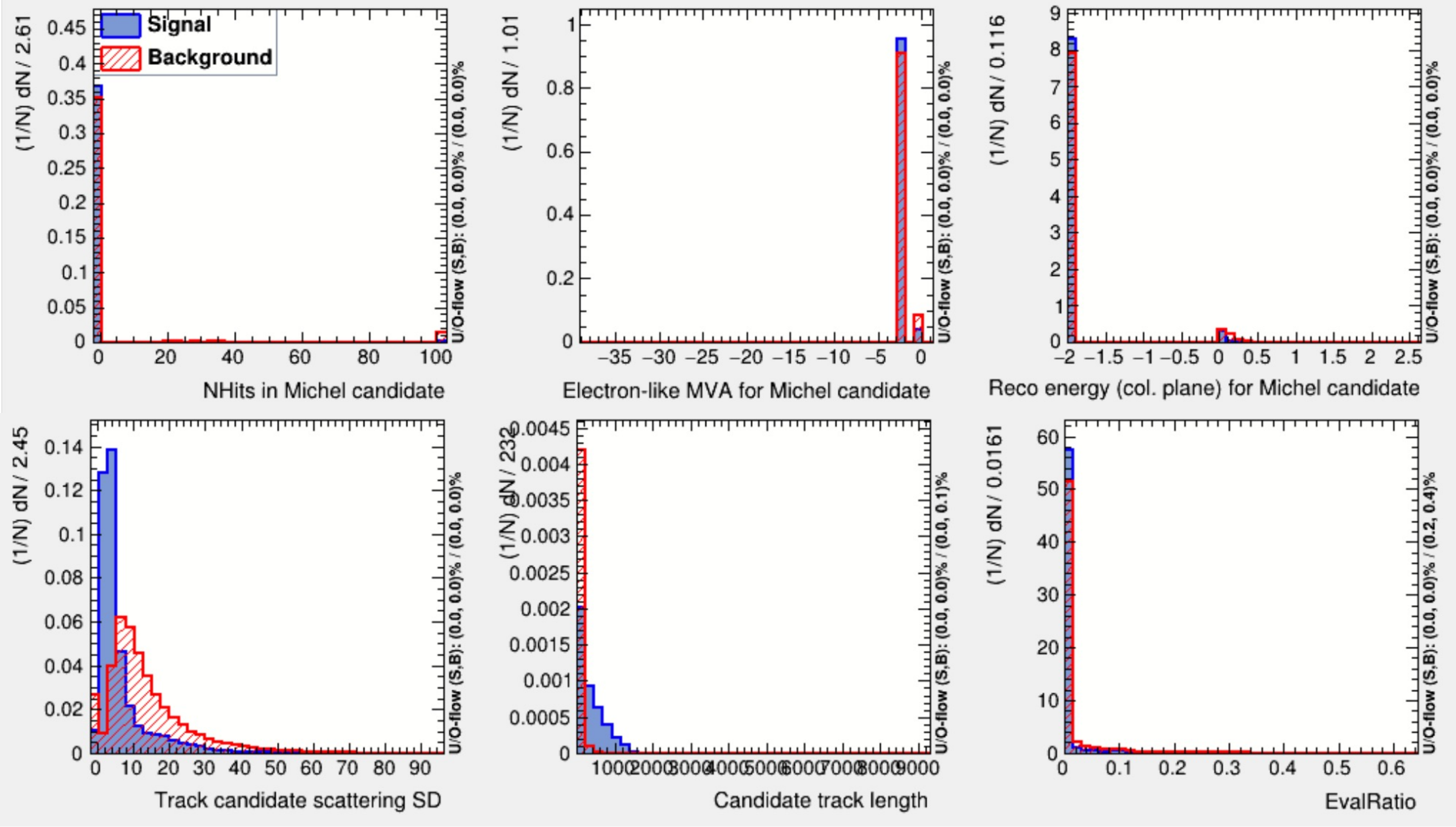
Pandizzle → muons
Pandrizzle → electrons

Tune cuts such that the selection **efficiency * purity** is optimised

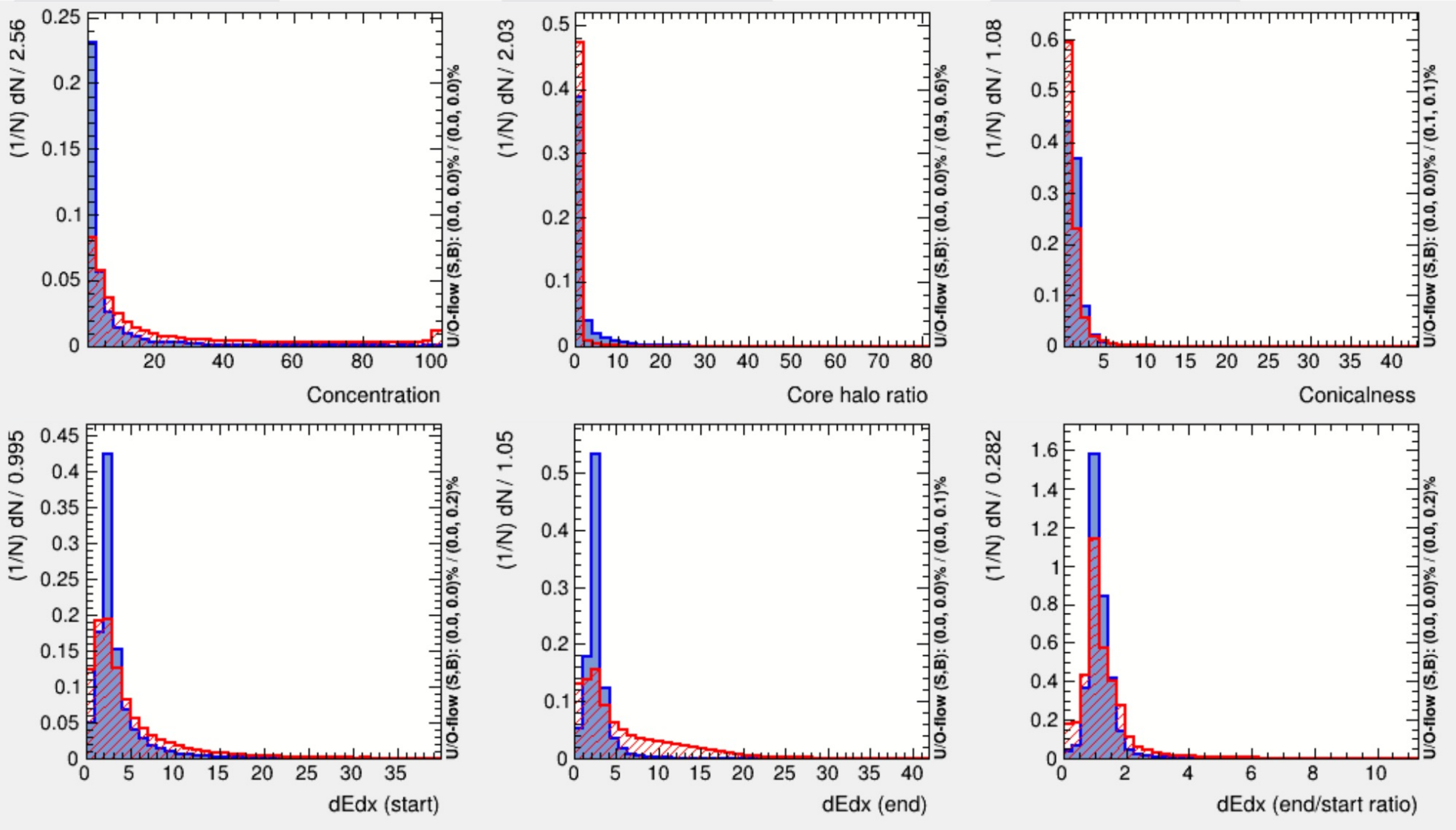
Pandizzle: ROC Curve



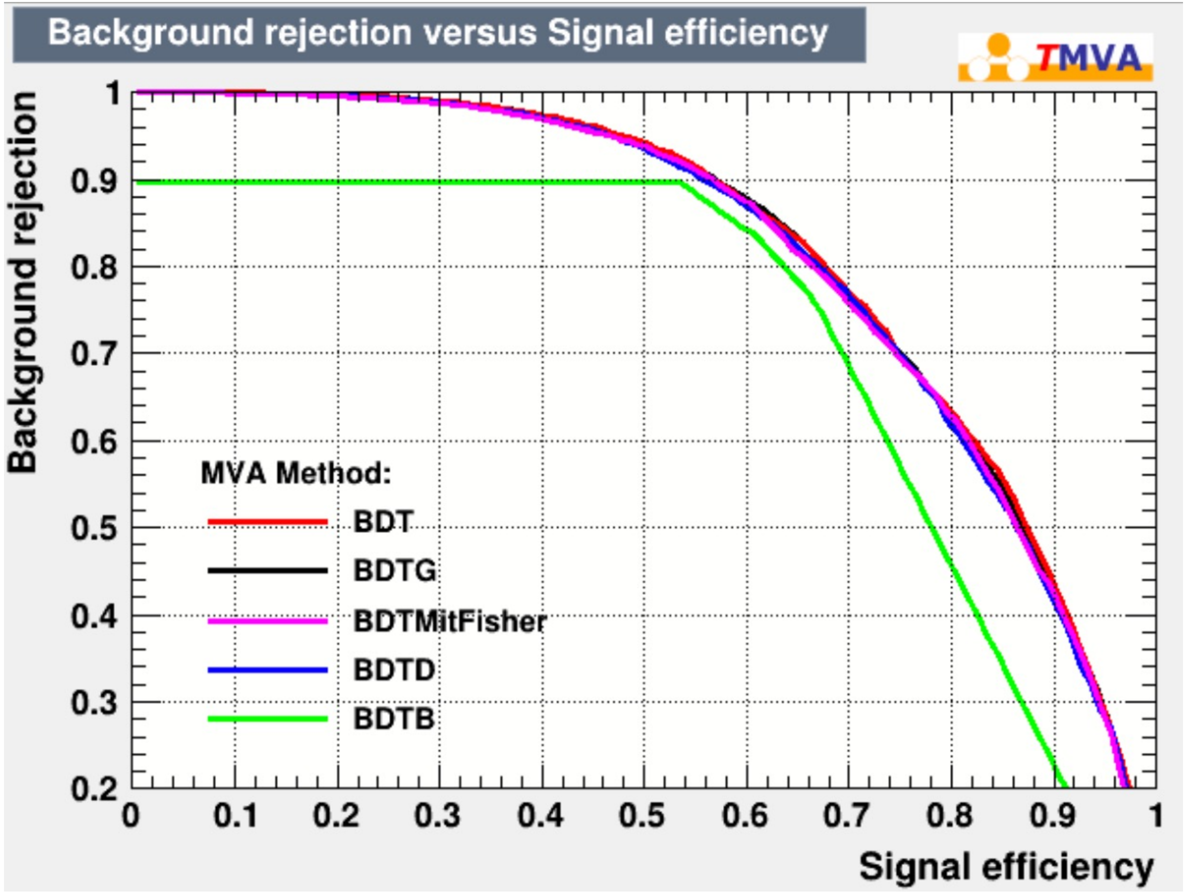
Pandizzle: Input Variables (1)



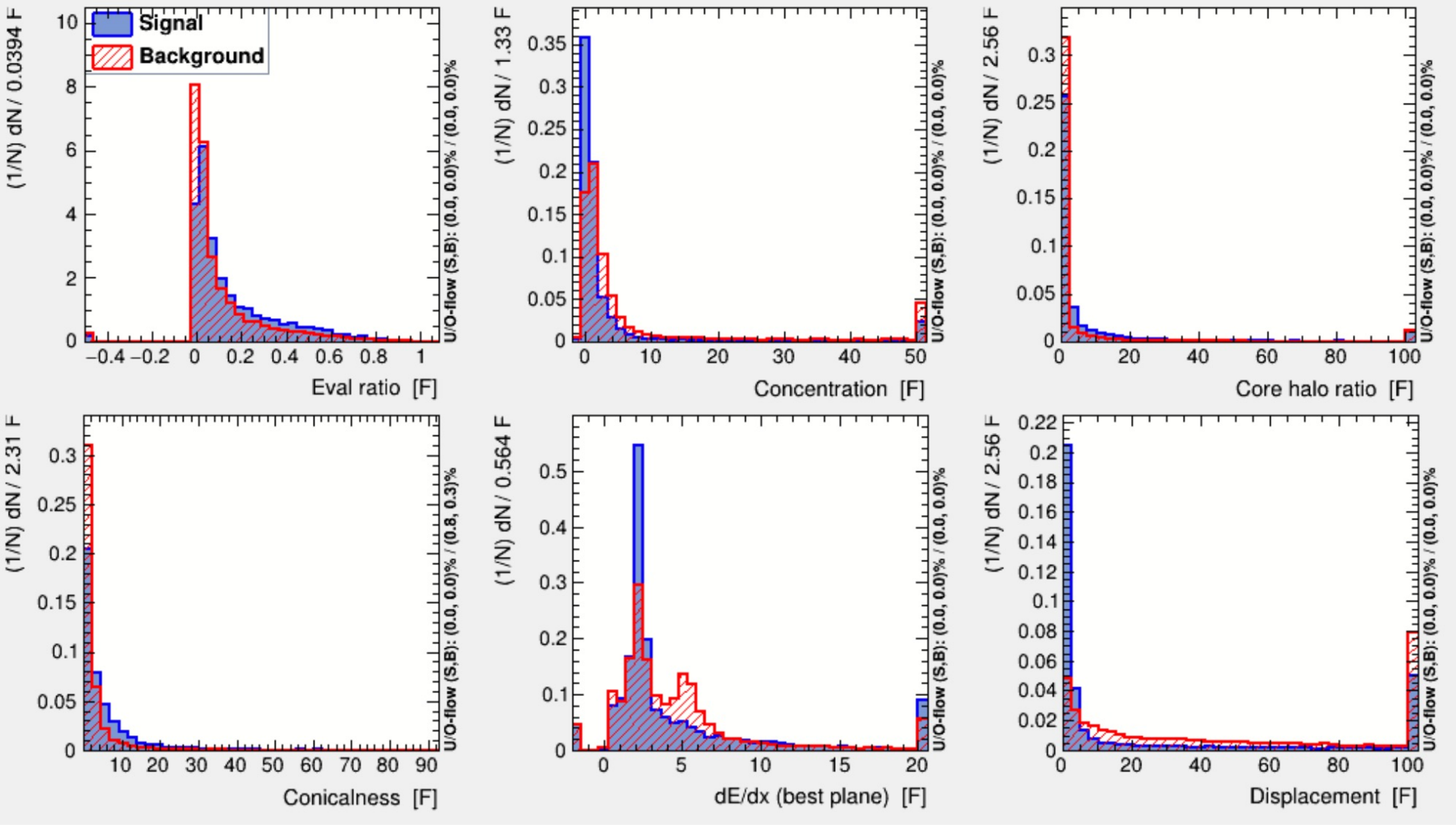
Pandizzle: Input Variables (2)



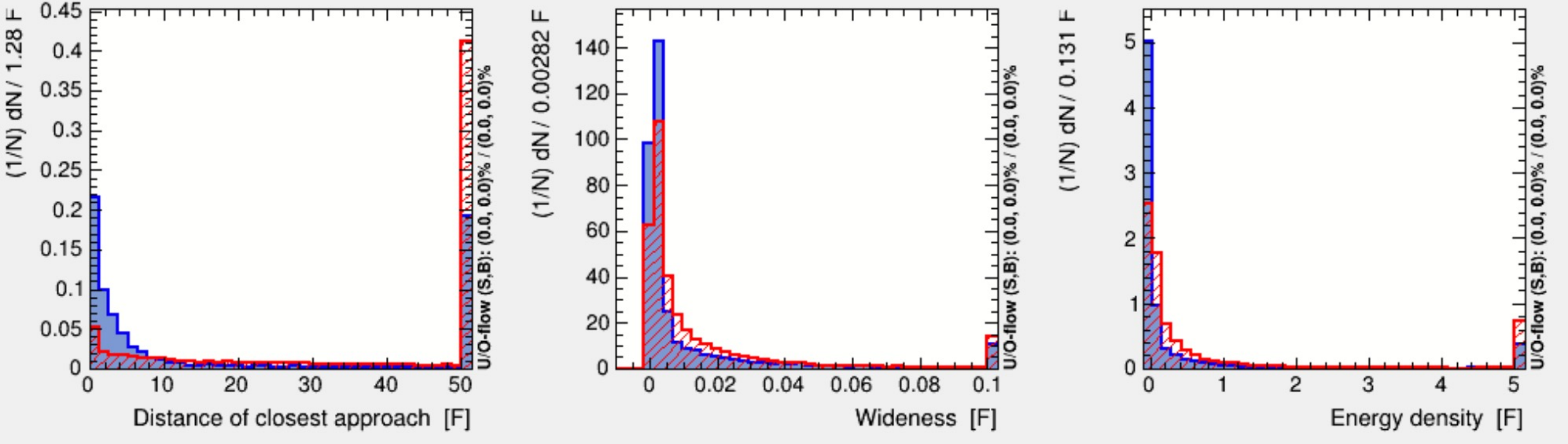
Pandrizzle: ROC Curve



Pandrizzle: Input Variables (1)

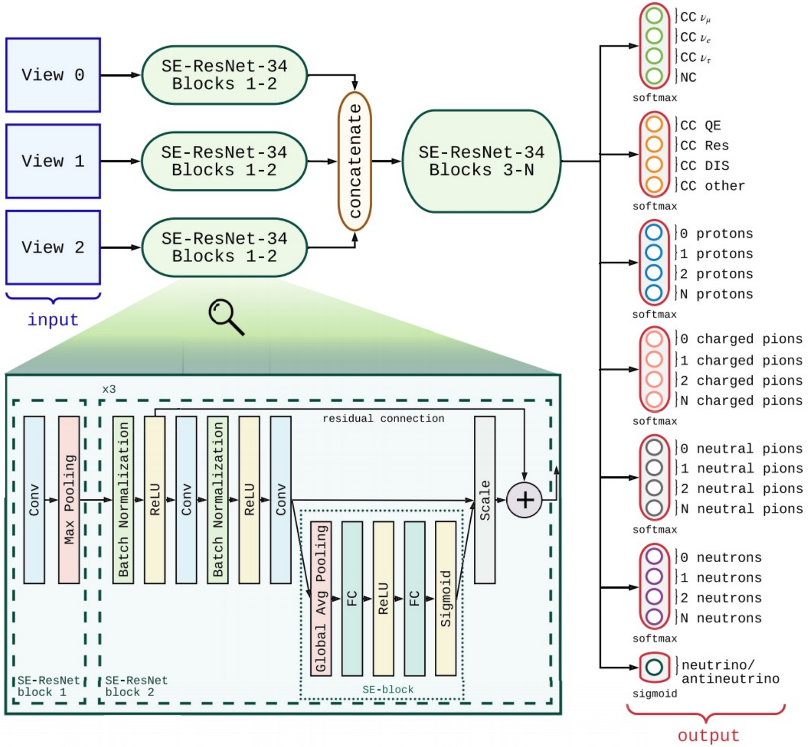


Pandrizzle: Input Variables (2)



DUNE CVN

- Details documented in https://journals.aps.org/prd/pdf/10.1103/PhysRevD.102.092003



- The DUNE CVN is a **convolutional neural network**
- Has access to the **whole** image
- For (a)nue selection: P(nue) > 0.85
- For (a)numu selection: P(numu) > 0.5

Tuning the nue cuts

- A 2D histogram is filled with the pandizzle and pandrizzle scores of all events with a reconstructed inside the DUNE fiducial volume
- Apply a test pandizzle and pandrizzle cut position to obtain the selection sample at all deltaCP values and consequently the corresponding deltaCP sensitivity plot
- Investigate the entire pandizzle-pandrizzle phase space, choosing the cuts that optimise the **deltaCP sensitivity coverage**